

10 Lecture 10

10.1 Characteristic polynomial

Lemma 10.1. Let $T : V \rightarrow V$ be a linear transformation, $v \in V$ an eigenvector of T with eigenvalue λ i.e. $Tv = \lambda v$. Let β, γ be two bases of V

$$\begin{array}{ccc}
 & & A[v]_\beta = \lambda[v]_\beta \\
 & \nearrow^{A=[T]_\beta^\beta} & \uparrow B = P^{-1}AP \\
 Tv = \lambda v & & [v]_\beta = P[v]_\gamma \\
 & \searrow_{B=[T]_\gamma^\gamma} & \downarrow \\
 & & B[v]_\gamma = \lambda[v]_\gamma
 \end{array}$$

We have $p_B(x) = p_A(x)$. We can define $p_T(x) = p_A(x)$ where $A = [T]_\beta^\beta$.

Proof. If $Tv = \lambda v$, then $[T]_\beta^\beta[v]_\beta = [T(v)]_\beta = \lambda[v]_\beta$.

$$p_{PAP^{-1}}(x) = \det(xI - P^{-1}AP) = \det(P^{-1}(xI - A)P) = \det(xI - A) = p_A(x). \quad \square$$

Proposition 10.2. The characteristic polynomial is of the form

$$p_A(x) = x^n - c_1(A)x^{n-1} + \cdots + (-1)^n c_n(A)$$

where $c_k(A)$ is some function of A for example, $c_1(A) = \text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$ and $c_n(A) = \det(A)$.

For any $P \in M_{n \times n}(F)$ invertible, we have $p_A(x) = p_{PAP^{-1}}(x)$. In particular $c_k(P^{-1}AP) = c_k(A)$.

If $p_A(x) = (x - \lambda_1) \cdots (x - \lambda_n)$ splits in F , then $c_k(A) = \sum_{i_1 < \cdots < i_k} \lambda_{i_1} \cdots \lambda_{i_k}$.

Proof. Note that $(xI - A)_{ij} = x\delta_{ij} - a_{ij}$. By the Leibniz formula of the determinant (Theorem 8.1)

$$p_A(x) = \det(xI - A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) (x\delta_{\sigma(1)1} - a_{\sigma(1)1}) \cdots (x\delta_{\sigma(n)n} - a_{\sigma(n)n}).$$

It is clear that $(-1)^n c_n(A) = p_A(0) = \det(-A) = (-1)^n \det(A)$. Therefore, $c_n(A) = \det(A)$.

To find the coefficient of x^n , we observe that the right hand side is the sum of products of n factors of degree ≤ 1 . The only term that can contribute x^n is the one where there is an x in each factor. This is only possible when $\sigma(i) = i$ for all $1 \leq i \leq n$, i.e. $\sigma = \text{id}$. Note that $\text{sign}(\text{id}) = 1$ and so this term is $(x - a_{11}) \cdots (x - a_{nn})$. Thus we see that the coefficient of x^n is 1.

To find the coefficient of x^{n-1} , we observe that the term that contribute x^{n-1} should have an x in at least $n - 1$ factors. This means $\sigma(i) = i$ for all but one $1 \leq i \leq n$. However, since $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ is bijective, we must have $\sigma(i) = i$ for all

$1 \leq i \leq n$. The term that contributes x^{n-1} is $(x - a_{11}) \cdots (x - a_{nn})$ again. The coefficient of x^{n-1} is $-a_{11} - \cdots - a_{nn} = -\text{tr}(A)$.

Suppose $p_A(x) = (x - \lambda_1) \cdots (x - \lambda_n)$. We now expand the product. To obtain the coefficient of x^{n-k} , we must choose the term x from exactly $(n - k)$ of the factors, and choose the term $-\lambda_i$ from the remaining k factors. Any such choice contributes $(-1)^k \lambda_{i_1} \cdots \lambda_{i_k}$. Since each k -tuple of distinct indices $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ occurs exactly once in the expansion, the coefficient of x^{n-k} is

$$(-1)^k \sum_{i_1 < \cdots < i_k} \lambda_{i_1} \cdots \lambda_{i_k}.$$

Comparing coefficients with $x^n - c_1(A)x^{n-1} + \cdots + (-1)^n c_n(A)$ we get

$$c_k(A) = \sum_{i_1 < \cdots < i_k} \lambda_{i_1} \cdots \lambda_{i_k}.$$

□

Example 10.3. $A = \begin{pmatrix} 7 & -10 \\ 5 & -8 \end{pmatrix}$, $\text{tr}(A) = 7 - 8 = -1$, $\det(A) = 7(-8) - 5(-10) = -6$. Thus $p_A(x) = x^2 + x - 6$. On the other hand, $P^{-1}AP = \text{diag}(-3, 2)$ and $\text{tr}(A) = -3 + 2$, $\det(A) = (-3)(2) = -6$.

Remark 10.4. There is a formula for $c_k(A)$ in terms of the entries of A which is the sum of principal k -minors of A . For a definition of principal k -minors and a proof of the formula, see here.

Corollary 10.5. $A \in M_{n \times n}(F)$ has at most n eigenvalues counting multiplicity. In other words, $(\lambda I - A)$ is invertible for all but at most n values of $\lambda \in F$.

Proposition 10.6. We have the following properties of the trace map.

- (i) $\text{tr} : M_{n \times n}(F) \rightarrow F$ is a linear function.
- (ii) Let $A, B \in M_{n \times n}(F)$. Then we have $\text{tr}(AB) = \text{tr}(BA)$.

Remark 10.7. (ii) is stronger than $\text{tr}(P^{-1}AP) = \text{tr}(A)$.

Remark 10.8. If $A, B, C \in M_{n \times n}(F)$, then by (ii) we have $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$. However, in general $\text{tr}(ABC) \neq \text{tr}(BAC)$.

Proof. (i) This is clear from definition.

(ii) Let $A = (a_{ij})$ and $B = (b_{ij})$. By the definition of matrix multiplication,

$$(AB)_{ii} = \sum_{k=1}^n a_{ik} b_{ki}.$$

Therefore,

$$\text{tr}(AB) = \sum_{i=1}^n (AB)_{ii} = \sum_{i=1}^n \sum_{k=1}^n a_{ik} b_{ki}.$$

We may interchange the order of summation:

$$\operatorname{tr}(AB) = \sum_{k=1}^n \sum_{i=1}^n a_{ik} b_{ki} = \sum_{k=1}^n \sum_{i=1}^n b_{ki} a_{ik}.$$

But this is exactly:

$$\operatorname{tr}(BA) = \sum_{k=1}^n (BA)_{kk} = \sum_{k=1}^n \sum_{i=1}^n b_{ki} a_{ik}.$$

Hence, $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. □

10.2 Diagonalizability

Theorem 10.9. *Let $\lambda_1, \dots, \lambda_m$ be distinct eigenvalues of $A \in M_{n \times n}(F)$. Let v_i be an eigenvector corresponding to λ_i for $1 \leq i \leq m$. Then v_1, \dots, v_m are linearly independent.*

In particular $E(\lambda_1, A) + \dots + E(\lambda_m, A) = E(\lambda_1, A) \oplus \dots \oplus E(\lambda_m, A)$ and if β_i is a basis of $E(\lambda_i, A)$, then $\beta = \beta_1 \cup \dots \cup \beta_m$ is a basis of $E(\lambda_1, A) \oplus \dots \oplus E(\lambda_m, A)$.

Proof. We prove by induction on m .

If $m = 1$, then $v_1 \neq 0$. Hence v_1 is linearly independent.

Suppose the conclusion is true for $m - 1$, we now prove it for m . Let $a_1, \dots, a_m \in F$ be such that

$$a_1 v_1 + \dots + a_m v_m = 0. \quad (2)$$

Applying T on both sides and using they are eigenvalues, we get

$$a_1 \lambda_1 v_1 + \dots + a_m \lambda_m v_m = 0. \quad (3)$$

Multiplying λ_m on both sides of (2) we have

$$a_1 \lambda_m v_1 + \dots + a_m \lambda_m v_m = 0. \quad (4)$$

Subtracting (4) from (3), we get

$$a_1(\lambda_1 - \lambda_m)v_1 + \dots + a_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

By induction hypothesis, v_1, \dots, v_{m-1} are linearly independent. Thus $a_i(\lambda_i - \lambda_m) = 0$ for $1 \leq i \leq m - 1$. Since λ_i are distinct, we have $a_i = 0$ for $1 \leq i \leq m - 1$. Plugging back into (2), we get $a_m v_m = 0$. Since $v_m \neq 0$, we have $a_m = 0$.

We now show that $E(\lambda_1, A) + \dots + E(\lambda_m, A) = E(\lambda_1, A) \oplus \dots \oplus E(\lambda_m, A)$. If this is not true, then by Proposition 3.5, there are $w_i \in E(\lambda_i, A)$ for $1 \leq i \leq m$ with some $w_i \neq 0$ such that $0 = w_1 + \dots + w_m$. Let w_1, \dots, w_r be all nonzero w_i 's. Then $w_1 + \dots + w_r = 0$. This contradicts the fact that w_1, \dots, w_r are linearly independent.

Let $\beta_i = (v_1^{(i)}, \dots, v_{k_i}^{(i)})$ be a basis of $E(\lambda_i, A)$. Then we would like to show that $\beta = (v_1^{(1)}, \dots, v_{k_1}^{(1)}, \dots, v_1^{(m)}, \dots, v_{k_m}^{(m)})$ are linearly independent. Let $a_l^{(i)} \in F$ be such that

$$\sum_{i=1}^m \sum_{l=1}^{k_i} a_l^{(i)} v_l^{(i)} = 0.$$

Then with $w_i = \sum_{l=1}^{k_i} a_l^{(i)} v_l^{(i)}$, we have $w_1 + \cdots + w_m = 0$. Since $E(\lambda_1, A) \oplus \cdots \oplus E(\lambda_m, A)$ is a direct sum, $w_i = \sum_{l=1}^{k_i} a_l^{(i)} v_l^{(i)} = 0$. Since β_i is a basis, we have $a_l^{(i)} = 0$ for any $1 \leq i \leq m$, $1 \leq l \leq k_i$. \square

Proposition 10.10. *Let $A \in M_{n \times n}(F)$ and $\lambda \in F$ be an eigenvalue of A . Then we have the geometric multiplicity of $\lambda \leq$ the algebraic multiplicity of λ .*

Proof. Since λ is an eigenvalue of A , we have the geometric multiplicity $m = \dim E(\lambda, A) \geq 1$. Let $v_1, \dots, v_m \in F^n$ be a basis of $E(\lambda, A)$. We extend v_1, \dots, v_m to a basis $\beta = (v_1, \dots, v_n)$ of $V = F^n$. Let P be the matrix whose columns are v_1, \dots, v_n since $v_i \in F^n$ are column vectors. (Or one can say P is the change of basis matrix from the standard basis to β .) We would like to find $P^{-1}AP$. Since $Av_i = \lambda v_i$ for $i = 1, \dots, m$ and Av_j is unknown for $m+1 \leq j \leq n$, we have

$$P^{-1}AP = \begin{pmatrix} \lambda I_m & B \\ 0 & C \end{pmatrix}.$$

Then by Proposition 9.3

$$\begin{aligned} p_A(x) &= p_{P^{-1}AP}(x) = \det \begin{pmatrix} xI_m - \lambda I_m & -B \\ 0 & xI_{n-m} - C \end{pmatrix} \\ &= \det((x - \lambda)I_m) \det(xI_{n-m} - C) = (x - \lambda)^m p_C(x). \end{aligned}$$

In particular, $(x - \lambda)^m \mid p_A(x)$. By definition of algebraic multiplicity, $m \leq$ algebraic multiplicity of λ . \square

Theorem 10.11. *Let $A \in M_{n \times n}(F)$. Then the following are equivalent.*

- (i) A is diagonalizable over F
- (ii) p_A splits over F and geometric multiplicity = algebraic multiplicity for each eigenvalue.
- (iii) $n = \dim E(\lambda_1, A) + \cdots + \dim E(\lambda_m, A)$
- (iv) $F^n = E(\lambda_1, A) \oplus \cdots \oplus E(\lambda_m, A)$

Proof. (i) \implies (ii) If A is diagonalizable, then $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$. Then $p_A(x) = p_{P^{-1}AP}(x) = \det(xI - \text{diag}(\lambda_1, \dots, \lambda_n)) = (x - \lambda_1) \cdots (x - \lambda_n)$. For each λ_i , the algebraic multiplicity of λ_i is the number of times k_i that λ_i appear in $\text{diag}(\lambda_1, \dots, \lambda_n)$. The corresponding columns of P are linearly independent eigenvectors of eigenvalue λ_i . Thus the algebraic multiplicity k_i is equal to $\dim E(\lambda_i, A)$ which is the geometric multiplicity.

(ii) \implies (iii) Since $p_A(x)$ splits in F , we have

$$p_A(x) = (x - \lambda_1)^{k_1} (x - \lambda_2)^{k_2} \cdots (x - \lambda_m)^{k_m}$$

where λ_i are distinct roots of $p_A(x)$. Since geometric multiplicity = algebraic multiplicity, we have $\dim E(\lambda_i, A) = k_i$. Since $\deg p_A = n$, we have $n = k_1 + \cdots + k_m = \dim E(\lambda_1, A) + \cdots + \dim E(\lambda_m, A)$.

(iii) \implies (iv) By Theorem 10.9, (iii), Proposition 3.5 and Corollary 2.17, we have $F^n = E(\lambda_1, A) \oplus \cdots \oplus E(\lambda_m, A)$.

(iv) \implies (i) Let β_i be a basis of $E(\lambda_i, A)$. By Theorem 10.9, $\beta = \beta_1 \cup \cdots \cup \beta_m$ is a basis of $E(\lambda_1, A) \oplus \cdots \oplus E(\lambda_m, A) = F^n$. Thus we have a basis consisting of eigenvectors. \square

Corollary 10.12. *Let $A \in M_{n \times n}(F)$. If A has n distinct eigenvalues in F , then A is diagonalizable.*

Proof. For any eigenvalue λ_i , we have $\dim E(\lambda_i, A) \geq 1$. Then we have

$$\begin{aligned} n &\geq \dim(E(\lambda_1, A) + \cdots + E(\lambda_n, A)) \\ &= \dim(E(\lambda_1, A) \oplus \cdots \oplus E(\lambda_n, A)) \\ &= \dim E(\lambda_1, A) + \cdots + \dim E(\lambda_n, A) \geq n. \end{aligned}$$

Thus equality holds and we have A is diagonalizable. \square

10.3 Cayley-Hamilton theorem

Definition 10.13. Let $T : V \rightarrow V$ be a linear transformation on a vector space over F . Let $p(x) = a_n x^n + \cdots + a_0 \in F[x]$. Then we define

$$f(T) = a_n T^n + a_{n-1} T^{n-1} + \cdots + a_1 T + a_0 \text{id}.$$

Note that $f(T) : V \rightarrow V$ is also a linear transformation. Similarly, for $A \in M_{n \times n}(F)$, we define $f(A) \in M_{n \times n}(F)$ as

$$f(A) = a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 I.$$

Proposition 10.14. *Let $A \in M_{n \times n}(F)$ be fixed. The map $F[x] \rightarrow M_{n \times n}(F)$ $f(x) \mapsto f(A)$ preserves polynomial addition and multiplication. That is for $f(x), g(x)$ in $F[x]$ we have $(f + g)(A) = f(A) + g(A)$ and $(fg)(A) = f(A)g(A)$.*

Proof. Suppose $f(x) = \sum_{j=0}^m a_j x^j$ and $g(x) = \sum_{k=0}^m b_k x^k$. Then

$$(f + g)(x) = \sum_{j=0}^m (a_j + b_j) x^j \quad (fg)(x) = \sum_{j=0}^m \sum_{k=0}^m a_j b_k x^{j+k}.$$

Thus

$$\begin{aligned} (f + g)(A) &= \sum_{j=0}^m (a_j + b_j) A^j = \sum_{j=0}^m a_j A^j + \sum_{j=0}^m b_j A^j = f(A) + g(A) \\ (fg)(A) &= \sum_{j=0}^m \sum_{k=0}^n a_j b_k A^{j+k} = \left(\sum_{j=0}^m a_j A^j \right) \left(\sum_{k=0}^n b_k A^k \right) = f(A)g(A). \end{aligned}$$

\square

Lemma 10.15. *If $B = P^{-1}AP$, then $f(B) = P^{-1}f(A)P$.*

Proof. $B^k = (P^{-1}AP) \cdots (P^{-1}AP) = P^{-1}A^kP$

$f(B)v = a_nB^n + \cdots + a_0I = a_nP^{-1}A^nP + \cdots + a_0I = P^{-1}(a_nA^n + \cdots + a_0I)P = P^{-1}f(A)P.$ \square